Memory, 2019

Do children use different forms of verbal rehearsal in serial picture recall tasks? A multi-method study

S. Poloczek[a,b]*, L. A. Henry[c], D. J. Messer[c,d] and G. Büttner [a]

[a]*Department of Psychology, Goethe University Frankfurt and Center for Research on Individual Development and Adaptive Education of Children at Risk (IDeA), Frankfurt am Main, Germany;* [b]*School of Psychological Science, University of Bristol, Bristol, UK;* [c]*Division of Language and Communication Science, City, University of London, London, UK;* [d]*Centre for Research in Education and Educational Technology, Open University, Milton Keynes, UK*

Sebastian Poloczek, Department of Psychology, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 6, 60629 Frankfurt am Main, Germany; poloczek@paed.psych.uni-frankfurt.de

Do children use different forms of verbal rehearsal in serial picture recall tasks? A multi-method study

Use of verbal rehearsal is a key issue in memory development. However, we still lack detailed and triangulated information about the early development and the circumstances in which different forms of rehearsal are used. To further understand significant factors that affect children's use of various forms of rehearsal, the present study involving 108 primary school children adopted a multi-method approach. It combined a carefully chosen word length effect method with a self-paced presentation time method to obtain behavioural indicators of verbal rehearsal. In addition, subsequent trial-by-trial self-reports were gathered. Word length effects in recall suggested that phonological recoding (converting images to names - a necessary precursor for rehearsal) took place, with evidence of more rehearsal among children with higher performance levels. According to self-paced presentation times, cumulative rehearsal was the dominant form of rehearsal only for children with higher spans on difficult trials. The combined results of self-paced times and word length effects in recall suggest that 'naming' as simple form of rehearsal was dominant for most children. Self-reports were in line with these conclusions. Additionally, children used a mixture of strategies with considerable intra-individual variability, yet strategy use was nevertheless linked to age as well as performance levels.

Keywords: immediate serial recall; verbal rehearsal; memory strategies; development; overlapping waves theory

1 Introduction

Background. Verbal rehearsal is a key construct in models of immediate memory such as the working memory model (e.g. Baddeley, 1986), and refers to the overt or covert repetition of to-be-remembered items. It is theorised that through articulatory or subvocal rehearsal, memory representations are maintained or refreshed to reduce forgetting. In many current conceptualizations, rehearsal is assumed to be one of the key maintenance mechanisms for storing verbal information (Camos, 2015; although see Lewandowsky & Oberauer, 2015). The present study adopted a novel multi-method approach to shed more light on verbal rehearsal in children between the ages of 6/7 and 10 years who, according to previous research (e.g., Flavell, Beach, & Chinsky, 1966), are likely to have started using some form of rehearsal. More specific evidence indicates that most 7-year-olds use simple verbal naming and it is only among 10-year-olds that the more developmentally advanced cumulative rehearsal is the most commonly observed classification

(Henry et al., 2000).

Overt rehearsal, as well as subvocal rehearsal, is assumed to rely on articulatory processes. In Baddeley's working memory model, rehearsal is conceptualized as the supportive mechanism underlying the refreshing of storage in verbal short-term memory. This storage is limited in capacity (e.g. up to two seconds, Baddeley, 1986); further, the number of items that can be rehearsed and retained is limited by the time needed to articulate the items (most often the syllabic length of the words) and the individual's articulation rate. Research over the past decades with adults has investigated word length effects in immediate serial recall performance, as well as correlations between speech rate and memory span (e.g., Baddeley, Lewis, & Vallar, 1984). This work supports the view that rehearsal can be used as a maintenance mechanism for verbal material, although it does not rule out conceptualising rehearsal as a voluntary strategy that may or may not be utilised.

Related work has examined important developmental progressions towards the development of rehearsal by children in immediate verbal serial recall tasks (e.g., Henry, Messer, Luger-Klein, & Crane, 2012; Hulme, Thomson, Muir, & Lawrence, 1984). For many years it has been assumed that children develop adult-like cumulative rehearsal, repeating more and more words, from around the age of seven years, and that this contributes to developmental increases in children's verbal short-term memory capacity, or 'memory span' (Flavell et al., 1966; Gathercole, 1998; Morra, 2015). However, Jarrold and Hall (2013) criticised the evidence put forward to support this qualitative change, arguing that several of the reported differences between younger and older children could be due to measurements being less reliable in younger children with lower memory spans, and word length as well as phonological similarity effects likely being proportional to span – therefore automatically smaller in younger children. So, the question of when and how children develop rehearsal is not settled yet. We next consider research investigating the development of rehearsal in children, including work from several areas: word length effects; self-paced recall; and self-reports.

Word length effects. Theoretically-driven research in the working memory tradition has often utilised word length effects to reveal more about underlying mechanisms of immediate serial recall. These effects, particularly when assessed using non-verbal presentation methods (e.g. easily nameable pictures), provide a useful indicator of whether or not children have developed verbal rehearsal, as pictures with longer names take longer to be verbalised. Previous research is not entirely in agreement over the age at which word length effects emerge with visual presentation of picture lists in memory span tasks (e.g. Henry et al., 2012; Henry, Turner, Smith & Leather, 2000; Hitch, Halliday, Dodd, & Littler, 1989). Further, even when word length effects are reported, the methodology cannot distinguish whether these stem from the process of phonological recoding, i.e. subvocally naming each picture after its presentation, or from processes involving cumulatively rehearsing several picture names together. Thus, when effects are reported in samples of young children, these are typically interpreted conservatively, i.e., only as evidence of a naming process (Henry et al., 2012). Therefore, it is important to revisit the question of the relationship between word length effects and the rehearsal processes that could be responsible for such effects.

Self-paced recall. A second approach to understanding the use of rehearsal has focussed on assessing the time participants take to inspect each picture item as it is presented (Belmont & Butterfield, 1971). This long-neglected method relies on allowing participants to determine a 'self-paced' presentation rate for each item in a serial list: if participants use cumulative rehearsal then longer inspection times are expected with progression through the list to allow more time for cumulative rehearsal. This method can also provide information about whether more time is taken with long rather than short-named items, due to the increased time burden of longer items during phonological recoding or related processes such as cumulative rehearsal. Belmont and Butterfield (1971) successfully used this technique with students, finding that typical adolescents paused longer as they went further into the to-be-remembered lists (although those with intellectual disabilities did not). More recently, Poloczek and colleagues (2016) used this method as part of a picture memory span word length experiment with children, which revealed that high span individuals with and without intellectual disabilities had longer inspection times for long than short-named pictures. An advantage of this method is that it allows a distinction to be made between simple verbal rehearsal involving naming each new item as it is presented, and cumulative rehearsal, whereby after each new item all previous items in the list are rehearsed together.

Self-reporting. Developmental studies in the working memory tradition typically treated rehearsal as a mechanism in working memory, comparing averages of experimental effects in older versus younger children, to determine when rehearsal as a feature in working memory develops. However, rehearsal does not have to be conceptualized as a hard-wired maintenance mechanism, rather verbal rehearsal could be a strategic behaviour that (some) children and adults (sometimes) choose to employ when asked to retain information. Investigating strategy use typically relies on trial-by-trial self-reports or 'think aloud' procedures, which are powerful methods of revealing detailed information about multiple-strategy use, but have been largely neglected in studies examining rehearsal development in serial recall (although see McGilly & Siegler, 1989). Self-reports and think aloud procedures are more commonly used in developmental research with free recall tasks. Previous research with free recall tasks demonstrates that children draw on multiple forms of rehearsal and other strategies when remembering lists of items and that the shift from naming to cumulative rehearsal is gradual over several years (e.g. Lehmann & Hasselhorn, 2007). Rehearsal in these studies is discussed in the context of adaptive strategy choice models. However, the development of rehearsal might differ between serial recall and free recall. This is because the pace of item presentation is typically faster in serial than free recall tasks, and the amount and form of rehearsal (single word vs cumulative), at least in undergraduates, strongly depends on the rate of presentation. For example, there is virtually no overt cumulative or partial cumulative rehearsal at presentation rates as rapid as one word per second (Tan & Ward, 2008).

Other factors. There are some further factors that may impact on the use of rehearsal. For example, the likelihood of using cumulative rehearsal could be linked to age as well as memory span level. Jarrold and Hall (2013) have argued that individuals may attempt to cumulatively rehearse items they can successfully remember, such that an individual's current rehearsal capacity (and potentially their choice of the form of rehearsal) is constrained by recall capacity (see also Cowan & Vergauwe, 2015). However, these authors have not yet supported this position with empirical evidence. Alternatively, the likelihood of using cumulative rehearsal could be adaptively adjusted to the difficulty of the memory trial. For example, children might be more encouraged to use cumulative rehearsal on challenging trials with slightly more words than their memory span, while on trials still within their memory span, simpler strategies such as verbal naming might appear sufficient for good recall. If cumulative rehearsal is an adaptively used strategy, we would expect an effect of task difficulty. If cumulative rehearsal is a hard-wired maintenance mechanism, no task difficulty effect would be expected.

The current study. Previous research has highlighted three different approaches to understanding verbal rehearsal in children. In the current study, we have combined them all to provide strong and methodologically robust descriptions of the use of different forms of rehearsal in serial recall tasks. Two methods rely on behavioural indicators: (1) measures of the word length effect in serial recall of visually presented nameable pictures; and (2) the analyses of presentation times children selected when given the freedom to self-pace through serial recall lists. These indicators have the advantage of not depending on participants' meta-cognitive abilities to report accurately on rehearsal. However, to infer rehearsal, data have to be aggregated across trials and across participants and information about the variability in rehearsal is hard to obtain. As the literature has demonstrated, for example on free recall, rehearsal use can vary inter-individually as well as intra-individually (e.g. Lehmann & Hasselhorn, 2007). Consequently, we decided to include a third measure, self-reports, to assess rehearsal use and other ways of remembering trial-by-trial.

Research questions. Two key questions were addressed using multiple methods to determine whether the findings obtained were consistent and led to common conclusions. As this has rarely been done in previous research on rehearsal development, the multi-method approach was an important novel feature of this study. The first question to be answered using triangulated information from the three methods was: what forms of rehearsal did children use to support recall? The hypothesis was that cumulative verbal rehearsal and naming (as a simple form of verbal rehearsal) would be the most common approaches based on previous research (e.g. Henry et al., 2000). The second question was: did task difficulty, age and baseline memory span influence the use of different forms of rehearsal? We hypothesized that cumulative rehearsal would be more common among older children and more common on the more challenging trials given evidence from the free recall literature that different forms of rehearsal can be conceptualized as adaptive strategies (Lehmann & Hasselhorn, 2007). Following Jarrold & Hall (2013) we hypothesised that children with higher spans would be more likely to use cumulative rehearsal because items need to be recalled to rehearse them. Disentangling these influences is important, as age (Henry et al., 2000) and memory span level (Lehmann & Hasselhorn, 2007) might affect the use of cumulative rehearsal.

2 Method

2.1 Participants

Children were recruited from 18 classes ranging from grade 1 (6-7 years) to grade 4 (9-10 years) in seven mainstream German primary schools. This range reflects the most sensitive ages for the development of verbal rehearsal: in particular, the shift from simple naming to cumulative rehearsal (Henry et al., 2000). To be included in the study children had: (a) to be without a special needs statement, as memory strategies can differ between children with intellectual disabilities and chronological age-matched typically developing peers (Henry, 2008); and (b) to have German as their first language to ensure the word length manipulation would be effective with verbal strategy use when remembering the presented pictures. Written informed consent from parents/guardians and verbal assent from children were additional inclusion requirements. Before conducting the study, ethical approval was obtained from the relevant University department ethics board.

Two of the 110 participating children, both aged 7 years, only obtained a picture span of 1 during baseline testing. They completed the word length experiments with trials consisting of one or two pictures. However, they were not included in the data analyses because, they did not have enough data to be included in presentation time analyses.

The final sample consisted of 108 children (50.0% female) between the ages of 6.4 years and 11.1 years with a mean age of 8.3 years. Age and picture span (see Section 2.2.3) were correlated (r = .33, p = .001), but there were also some children aged 6 or 7 years with spans of 4 and above, as well as children aged 9 or 10 with spans of 3 and below (see Table 1). Nonverbal IQ, measured with Raven's Standard Progressive Matrices (SPM, Raven & Horn, 2009), was in the range of 72 to 139 with a mean of 109.8 (SD = 16.4). Children with IQs in the borderline range (70-79, n = 3) but without a special needs statement were included to ensure that this sample was representative of mainstream children.

(Table 1 about here)

2.2 Procedure and Materials

2.2.1 Methodological considerations

The procedure chosen to collect information about word length effects involved the recall of *visually* presented *easily-nameable* items, with recall via pointing to pictures in serial order. As the input is visual, no word length effects are expected unless: (1) phonological (verbal) recoding of the picture images into their corresponding names had occurred; and (2) the child attempts to retain the list in a verbal format. Furthermore, this procedure avoids the possibility that merely verbal output is responsible for word length effects, simply because longer words take longer to produce during recall (Cowan et al., 1992; Henry, 1991). A proportion of any resulting word length effect can be generated during output, when pointing to the images is driven by subvocal articulation of the remembered words. However, as this can only occur if the visual information has been converted into a verbal code, this output contribution does not invalidate the inference to verbal strategy use.

We also included trial-by-trial self-reports of task behaviour to provide important new evidence over and above other markers of rehearsal that have to be aggregated and to gain inside into rehearsal use being strategic. Given that: (1) children's awareness of their own thoughts is less likely when these thoughts are spontaneous, undirected and unintended (Flavell, Green, & Flavell., 2000); and (2) we do not know how intentional children's use of naming or cumulative rehearsal is, we need to be careful when using self-reports on rehearsal behaviour. Flavell and colleagues have argued that valid introspections are more likely if children know in advance that they will be asked to report their cognitive strategies, if they are given sufficient practice, and are provided with good retrieval cues at the time of reporting. As these attempts to facilitate valid introspections could alter strategic behaviour, we introduced self-reports only on the second half of the remembering trials so that the results could be compared to the behavioural indicators obtained during the first half of trials which had no explicit strategy assessment.

2.2.2 Overview of testing sessions

Participants were tested at their schools in four sessions. Raven's Standard Progressive Matrices (SPM) was administered in groups of up to five students during the first session of approximately 30 min. This was a paper and pencil test; all other tasks were administered individually on a computer. During the second session of approximately 10 min, baseline picture memory span was determined. Sessions 1 and 2 took place on the same day while the remaining sessions were administered on different, non-consecutive days. The memory experiment was administered without explicit strategy assessment during the third session (approximately 25 min). Finally, during the fourth session of approximately 30 min, the memory experiment was administered with strategy assessment, five strategy options and after each trial children were asked to indicate the strategy they had used for that trial.

2.2.3 Picture sets

Each of the three picture/word sets used in the study consisted of nine items which were familiar, concrete, highly imageable objects with a low age of word acquisition. The set for assessing baseline picture span and the set for the short word condition in the word length experiment were composed of black-and-white line drawings with monosyllabic names like *book/Buch*, *dog/Hund*, *glass/Glas* and *skirt/Rock*. The set for the long word condition in the word length experiment was composed of black-and-white drawings with trisyllabic names like *pineapple/Ananas* and *telephone/Telefon*. For further details, including the complete word lists and ages of acquisition see Poloczek et al. (2016).

Before the baseline picture span task and the word length experiment, the experimenter showed all pictures relevant to the task and named each once to make sure that children heard the intended name and were more likely to use these labels if they wished to name the items. The experimenter did not encourage participants to name the pictures or to repeat the names to avoid inducing verbal rehearsal.

2.2.4 General procedure for memory trials

All pictures appeared in the centre of the screen. Half a second after the last picture was presented, a response slide was displayed consisting of all nine items from the picture set in a 3 x 3 array. Five different response slides with pictures in different random orders were used and they were selected randomly to prevent children from learning the spatial locations of pictures within the response array. Participants were asked to indicate the pictures they remembered by pointing to them in the correct serial order. Verbal responses were not accepted to avoid encouraging verbal rehearsal. All these methodological features (timing of presentation and recall, avoiding spatial cues during presentation and response, responding in correct serial order) were chosen to make the memory trials as similar as possible to a verbal serial recall task, with the key difference that verbal input and output were avoided.

2.2.5 Baseline picture span

Pictures were presented for 1.5s with 0.5s inter-stimulus intervals. Assessment of baseline picture span started with lists of one item, followed by longer lists until the participant had reached his/her span level. Up to six trials were presented at each list length, but if the first four trials were entirely correct, trials 5 and 6 were skipped and scored as correct. Whenever participants correctly remembered at least four out of six trials of a particular list length, testing continued with lists lengthened by one item. When three or more trials at a list length were incorrect, baseline testing was completed. Span was the longest list length for which more than half of the trials (at least four out of six trials) were repeated completely in correct order. Span+1 was defined as one above span level.

2.2.6 Word length experiment with fixed and self-paced presentations

The number of pictures presented per trial depended on the child's baseline performance to ensure that task difficulty was comparable between participants and was adequately titrated to ability levels, in order to support the use of cumulative rehearsal. If trials were too easy, rehearsal might not be necessary to ensure recall. If trials were too demanding, cumulative rehearsal might break down and not be used for later list items (Lehmann & Hasselhorn, 2007; Tan & Ward, 2008).

In the condition with a fixed presentation rate, the procedure was similar to the baseline task. All pictures were presented for 1.5s with 0.5s inter-stimulus intervals. To-be-recalled picture lists were presented one after another for six trials at span level and for a further six trials at span+1 level. Trials with short-named items alternated with trials with long-named items. Pictures recalled in the proper serial position were recorded as correct. The proportion of correctly remembered pictures was calculated for the descriptive tables, but the binary correct-incorrect values were the dependent variable for recall performance in the generalized linear mixed models.

In the condition with self-paced presentation, the participants determined the presentation time for each picture. The time between the start of the presentation of each picture and the next keystroke was logged in milliseconds and was used as dependent variable (presentation time). Several rules were used to ensure valid presentation times. Very short (<300ms) and very long (>15s) presentation times and all presentation times below or above 2.5 standard deviations from the individual's mean presentation time were set to missing, as these presentation times were interpreted as reflecting atypical responses.

The self-paced and fixed presentation conditions were identical in all other respects. Because we aimed to examine strategic behaviour in 'typical' immediate serial recall tasks, the fixed condition preceded the self-paced condition to prime a moderately fast-paced presentation rate, similar to that in serial recall tasks.

The procedure of the experiment with all its manipulations of word length, list length and pacing was exactly the same for session 3 without self-reports and session 4 with self-reports; only explaining the response options and asking for self-reports was added in session 4.

2.2.7 Self-reports

At the beginning of session 4, the experimenter explained that children do different things to remember the pictures. Five ways of remembering were introduced while the child was shown sequences of three illustrative pictures (see Figure 1): (1) naming each picture individually inside one's head or whispering each name (naming); (2) after the presentation of the very last picture, repeating all picture names to oneself (complete rehearsal); (3) as each new picture is presented, repeating all so-far presented pictures and always adding the new name to the list (cumulative rehearsal); (4) thinking of what a picture looks like or seeing the picture inside one's head (visual strategy); and (5) not doing anything special, just remembering the pictures (no strategy). Shorter versions of each strategy consisting of three thought bubbles were introduced after the initial explanations. Before starting with the memory trials, comprehension of the answering format was checked. Participants were told how different children remembered the pictures and were asked to indicate which thought bubbles each child would select. Then participants were told to indicate immediately after each trial which way they had remembered the presented pictures, by pointing to one of the five thought bubble sequences.

(Figure 1 about here)

2.3 Design

The tasks and procedure resulted in a design with three different dependent variables as rehearsal indicators. To test the word length effect on recall performance (the first rehearsal indicator), recall was assessed in the absence of any self-reports on 24 trials; the within-participant factors were word length (short vs. long words), difficulty (span vs. span+1 trials), and pacing (fixed vs. self-paced). The second rehearsal indicator, self-paced presentation times, was assessed on half of these 24 trials; the within-participant factors were word length and difficulty. As a third rehearsal indicator, self-reports were gathered on a further 24 trials; the within-participant factors were word length, difficulty and pacing.

2.4 Statistical Analyses

Initial descriptive statistics were done with SPSS 22. All further analyses were set up as (generalized) linear mixed models ((G)LMMs) and performed with MLwin 3.02 (Charlton, Rashbach, Browne, Healy & Cameron, 2018) with MCMC estimation (Browne, 2017) from R with the R2MLwin package (Zhang, Parker, Charlton, Leckie, & Browne, 2016). Using (G)LMMs provided a consistent, yet flexible modelling framework avoiding drawbacks of a more traditional ANOVA approach. For all our analyses (G)LMMs had the advantage, that dependant variables at the trial level can be modelled as depending on experimental conditions, person-level predictors and random person-level effects. Regarding the recall data and strategy choice data, analysing proportions correctly recalled or percentages of strategy chosen with ANOVAs can lead to spurious null results as well as spurious significance, a problem that can be avoided with logistic GLMMs (Jaeger, 2008). To analyse serial position effects in self-paced presentation times with a repeated measures ANOVA, all cases entered into the same analysis must have the same number of repeated measurements, i.e. presentation times for serial positions. This would result in multiple ANOVAs because difficulty (lists at span vs. lists at span+1) was experimentally manipulated and because list length was titrated to memory span and therefore varied between participants. In LMMs no such restriction to the same number of measurements applies. Therefore, data of all conditions and participants can be analysed within the same model and interactions with difficulty and memory span can be tested.

3 Results and Discussion

3.1 Rehearsal indicator 1 – word length effects in recall performance

To analyse the effect of the experimental conditions on the likelihood of recalling an item in correct serial position, a logistic GLMM with fixed effects for the experimental manipulations of trial difficulty or list length, word length, and pacing and the fixed effect for the participant level covariate of baseline memory span (participants with a span of 2 vs. 3 vs. 4 and higher) was set up. To mirror the approach of a repeated measures ANOVA, all interactions between these main effects were included as further fixed effects. Because trials were nested within participants, the random intercept for participants was included. The aims of the analyses were twofold: firstly, to examine whether there was a word length effect that would indicate phonological recoding and, therefore, verbal rehearsal and whether rehearsal varied with task difficulty; secondly, to test whether the word length effect varied with memory span.

3.1.1 The word length effect and task difficulty

A clear main effect of word length was found (see Tables 2 and 3), reflecting poorer performance on long-named pictures, and which is consistent with the use of verbal rehearsal. There was also a strong negative effect of list length; recall in the longer, more difficult span+1 trials was worse. There was no evidence for an interaction effect between word length and list length, as the coefficient for this interaction was close to 0 (with a narrow credible interval surrounding it).

(Tables 2 and 3 about here)

3.1.2 The word length effect and its interaction with children's baseline memory span

There was an unexpected, negative effect of span group or baseline memory span, due to a lower proportion of items recalled in correct position for children with higher spans. This negative effect of span group might be somewhat attenuated for trials at span+1, especially with long words, as there was some evidence for the list length x memory span interaction and the word length x list length x memory span interaction. However, these potential interactions have to be interpreted with caution, because the effect sizes were small in comparison to the main effect of memory span, the 95% credible intervals got very close to 0 and the Bayesian *p*-values were just below .05. The lower performance for children with higher spans was surprising, as the number of words per trial was titrated to the baseline span of each child. This suggests that it is somewhat harder for children with higher spans to do well in an experiment with span and span+1 trials. Omitting an early item to a serial list may disproportionately affect correct serial recall of subsequent items when the participant has a higher span. Alternatively, as span measures are not perfectly reliable there may have been biases. In particular, regression to the mean suggests that the actual span score of those with the lowest scores is more likely to be underestimated, while the scores of the highest performing children are more likely to be overestimated. Even though the difficulty level may not have been perfectly adjusted, our design choice has advantages over presenting all children with trials of the same list lengths. This previously used method of presenting the same number of words to all participants could have created ceiling effects by presenting easy trials for those with the highest spans; and would have created large differences in the likelihood of remembering items in correct position as well as disparities in task difficulty between low and high span individuals.

The interaction effect of word length and baseline memory span suggests that rehearsal might differ between high and low span children. Follow-up GLMMs showed, that there was no word length effect for children with a span of 2 (β = -0.09 [-0.35, 0.17], p = .50), and an increasingly larger word length effect emerged for individuals with a span of 3 (β = -0.31 [-0.40, -0.23], p < .001) and for those with a span of 4 or higher (β = -0.54 [-0.61, -0.46], p < .001). There is good evidence that the word length effect scales proportionately to memory span: the absolute difference between span for short and long words increases with the span of children, but the proportional cost or the difference between spans divided by span for short words is relatively constant (Jarrold, Danielsson, & Wang, 2015). Please note, that as list length was titrated to memory span, a form of adjustment to the proportional scaling of word length effects was included in the experimental design. Therefore, the increasing word length effect in children with higher baseline span levels despite titration to span level hints at more verbal rehearsal by individuals with higher spans.

3.1.3 The effect of pacing and its interactions

Additionally, an effect of fixed vs. self-pacing might have been present that was moderated by the interaction pacing x memory span (see Table 3). There was evidence that children with higher spans benefitted more compared to other children from being able to self-pace through the task. However, within each span group the recall difference between the fixed and the self-paced condition was small: with slightly poorer recall for low and medium span children in the self-paced condition and slightly better recall for this condition in children with higher spans (see Table 2).

In summary, word length effects on recall were observed that were larger for participants with higher baseline memory spans. This indicates the use of more verbal rehearsal by participants with higher spans. More verbal rehearsal could mean that rehearsal was used more often by children with higher spans, or that a more time-consuming form of verbal rehearsal like cumulative rehearsal, instead of mere naming or single word rehearsal, was used by these children, without a change in the number of trials in which pictures were phonologically recoded. To understand the nature of rehearsal use more fully, data from other sources than recall performance will be considered next.

3.2 Rehearsal indicator 2 – Self-paced presentation times

The patterns of self-paced presentation times (note this is not recall, rather it is how long the child chose to look at each item during the list presentation) were the second behavioural indicator of rehearsal. Analysing self-paced presentation times can address the following questions. Firstly, were there longer presentation times for pictures with long names, thereby suggesting that some form of rehearsal was used, and if so was there evidence that this effect was more prevalent in more challenging span+1 trials and in participants with higher baseline spans? Secondly, did presentation times increase with the serial position of items, i.e., allowing greater cumulative rehearsal time to accommodate increasing numbers of items? Such an effect would suggest that (some) participants were using cumulative rehearsal as an advanced form of rehearsal; the absence of this effect could suggest the use of a simpler form of rehearsal such as naming.

Mean presentation times for the self-paced condition of session without self-reports are provided in Table 4. Preliminary analyses and the inspection of the descriptive results revealed a large serial position effect from position 1 to position 2 with considerably longer presentation times for picture 1. This might have resulted from an artefact, the additional preparation time needed to press the key for the first time in each trial[1]. Therefore, the first position was excluded from analyses. To assess serial position effects, position was coded with the value 0 for position 2 and values increasing by 1 for all subsequent positions. Like in an ANOVA, all interaction effects were included. To take into account the clustering of presentation times within participants, the participant random intercept was allowed (improvement in the Deviance Information Criterion (Δ DIC) of 1753.0). The increasing standard deviations with serial position could be due to interindividual differences in the use of cumulative rehearsal. Therefore, the random slope for position was allowed and kept due to the improvement in model fit (Δ DIC = 212.6).

(Table 4 and Table 5 about here)

3.2.1 Word length effect and use of any form of rehearsal

No word length effect was found and all interaction effects with word length were close to 0 (see Table 5). Therefore, no evidence for phonological recoding was found. However, as the intercept indicates an average self-paced presentation of 1.89s, there was plenty of time to subvocally name pictures with short and long names. So, the lack of a word length effect on presentation times also cannot be taken as firm evidence against phonological recoding. It is noteworthy that children with higher spans allowed more time after each picture than those with lower spans.

3.2.2 Position effects and the presence of cumulative rehearsal

There might be a small negative effect of serial position, even after the first position with spuriously long RTs was excluded. But this effect was moderated by the interaction effects position x list length, position x memory span and position x list length x memory span. To interpret these position effects, follow up LMMs were run for children with medium or high spans[2]. For children with a span of 3, no main effect of position was present (β = -0.09 [-0.20, 0.02], *p* = 13), but a significant interaction of list length and position (β = 0.16, [0.08, 0.24], $p < .001$), as children sped up during trials at span level. In contrast for the group with a span of 4 and higher, was there a position effect (β = 0.13 [0.01, 0.27], p = 0.04) moderated by the interaction position x list length (β = 0.07 [0.01, 0.13], p = .02) due to a linear trend for increasing presentation times in later serial positions at span+1. This effect was consistent with the use of cumulative rehearsal at span+1.

The results can be summed up in the following way: support for cumulative verbal rehearsal was only found in the more difficult trials for children with high memory spans. For children with

---

[1] In a subsequent experiment, the procedure was changed slightly so that children had to press to get the first word presented and press again for the second word. When the difference between these two clicks was measured, presentation times for the first position were not longer. This suggests that longer times in the present study were a methodological artefact and did not mirror strategic behaviour.

[2] No follow up model was run for children with a span of 2, because this subsample was small and only few data points per child were available to estimate position slopes.

medium memory spans, self-presentation times did not show a linear trend for increasing times at later serial positions, suggesting their use of naming as a simple form of rehearsal or no form of rehearsal use.

3.3 Rehearsal indicator 3 – Self-Reports

While the behavioural indicators relied on recall or presentation time data being aggregated over trials and across participants, self-reports offer direct trial-by-trial information. Therefore, within-participant and between-participant variation in self-reported task behaviour can be examined. First, within and between participant variation is described. Second, GLMM analyses were conducted to investigate whether self-reports of task behaviour vary between experimental conditions. Additionally, the GLMMs examining the effects of participant characteristics in explaining between-participant variation in self-reports can help to disentangle potential effects of span, age, and IQ on task behaviour.

3.3.1 Descriptive results

Only 17.6% of the children reported always using the same way to remember, while 34.3% reported using each of the five ways in at least one trial. The remaining half of the children reported two (15.7%), three (11.1%) or four (21.3%) different ways, suggesting that task behaviour was varied, not hard-wired, and the different methods can be classified as optional strategies. More detailed information about reported strategies can be found in Table 6.

(Table 6 about here)

The descriptive results on self-reported task behaviour can be summarised as follows. There was considerable *between-participant* variation in how frequently a certain strategy was reported and which strategy was reported most often. Additionally, there was considerable *within-participant* variation as two out of three participants reported using three or more strategies. Additional descriptive results in Table 7 suggest that naming was the preferred strategy for children with spans of 2 or 3, and cumulative rehearsal was the preferred strategy for children with higher spans (>=4).

(Table 7 about here)

3.3.2 Dependence of strategy use on experimental conditions and person characteristics

To test whether within-participant variation in strategy use can be (partially) explained by the experimental conditions and whether between-participant variation was significantly linked to participant characteristics, GLMMs were run with experimental conditions as trial-level predictors (level 1) and baseline memory span, age, nonverbal IQ, and gender as person-level predictors (level 2).

Table 8 provides the results of the five final GLMMs predicting whether a certain strategy was reported or not. Including random intercepts improved model fit (clear change in DIC) suggesting that for all strategies there was still significant between-participant variation in the likelihood that certain strategies were used even after taking all child level predictors into account.

(Table 8 about here)

Level 1 predictors or dependence on task conditions

In the model for naming, the negative coefficient for difficulty indicates that on the more difficult span+1 trials compared to span trials, naming was less likely (β = -0.21***, 39.5% for span vs. 34.1% for span+1[3]) while cumulative rehearsal was more likely (β = 0.25***, 31.2% for span vs. 37.3% for span+1). The coefficients for list length were close to 0 in the models for the other strategies. This suggests that complete rehearsal, visual or no strategy use was equally reported in easier and harder trials.

Word length was not a relevant variable in any of the models predicting strategy reports, the coefficients for word length were all close to 0. It is important to note that in these analyses (in contrast to the ones in 3.1 and 3.2) a word length effect was not expected as an indicator of verbal strategy use. The null effect of word length indicates that children did not alter (their self-reported) strategic behaviour for pictures with short vs. long names. The coefficients for pacing were, similarly, close to 0, although in the case of the model for cumulative rehearsal there might be a small effect. The estimate (β = 0.10, p = .07^{+}) could indicate that cumulative rehearsal was slightly more prevalent in trials with self-paced compared to fixed presentation (33.0% vs. 35.4%), but any effect was small.

Level 2 predictors of inter-individual differences

In all models, model fit was improved when by-participant random intercepts were introduced, indicating that children reliably differed in their strategy preferences. Therefore, level 2 predictors were introduced to examine which participant characteristics explained variance in children's preferences for certain strategies. No effect of gender was found in any of the models. Also all 95% credible intervals for the predictor contrasting children with a span of 2 with those in the reference category with a span of 3 included 0, so no evidence for a difference in strategy choices between those with low and medium spans were found. However, as there were only 13 children with a span of 2, credible intervals were quite large so this result needs to be interpreted with caution. It might be that children with low spans were more likely to report a visual strategy or no strategy, but that the subsample was too small to provide evidence for such a difference.

Children with high spans (>= 4) were less likely to report naming (23.2% of the trials vs. 46.7% of the trials in the reference group with a span of 3) and more likely to report cumulative rehearsal (59.0% vs. 21.2%). The clearly negative coefficient of -1.94*** in the model for naming and the large positive value of 2.16*** in the model for cumulative rehearsal indicate a strong impact of belonging to the group with a high span vs. the reference group on the choice of strategy. As age and span were correlated (r = .33, p = .001) it was of particular interest to include both age and span to test whether one or both of these predictors were linked to strategy use. The estimate of 0.72*** for age in the model predicting cumulative rehearsal indicated that older children were more likely to report cumulative rehearsal. Therefore, both memory span group and age contributed to predicting the likelihood of using cumulative rehearsal. The change in the likelihood of using cumulative rehearsal that was associated with changing from the reference group to the group with a high span (when age was held constant) was as large as the change associated with increasing the age by about

---

[3] These probabilities refer to the percentage of trials in which the strategy was reported aggregated across all children. Percentages are provided for illustration as coefficients in logistic models can be unintuitive.

3 years when span was held constant. Age was not linked to reporting naming as strategy, but it was a negative predictor in the model for complete rehearsal (i.e. the simpler 'repeating the full list once' rehearsal). Therefore, older children reported less complete rehearsal.

Nonverbal IQ was also linked to the strategy preferences. When differences in span were controlled children with higher nonverbal IQ were more likely to report naming (β = 0.71*** for an IQ change of 1 *SD*) and less likely to report complete rehearsal (β = -0.57***) or a visual strategy (β = -0.41*).

4 General Discussion

The results from all three methods converge on the conclusion that primary school age children mainly, but not exclusively, used forms of verbal rehearsal to remember pictures in an immediate serial recall task. The dominant self-reported strategies were: (1) naming each picture; and (2) cumulative rehearsal of the already presented pictures. Naming was the most frequent strategy for children with low to medium baseline memory spans; and cumulative rehearsal was the most frequent for those with high spans. For children with medium and high baseline spans, these conclusions were consistent across all three rehearsal indicators.

For children with a span of 2, naming was the most frequently reported strategy. However, for this subgroup, phonological recoding was not corroborated by significant word length effects either for recall or for presentation times. This could mean that these children did not report accurately their actual task behaviour. Alternatively, it is possible that the small sample of these children (13) and there being fewer words per trial because of the design resulted in lower power to detect differences. Differences in presentation time due to word length might also be hard to detect, given that among primary school children the speech rate for one-syllable nouns is about 3 words per second and for three-syllable nouns about 1.6 words per second (Ferguson, Bowey, & Tilley, 2002), most self-presentation times in the current study were between 1 and 3 seconds per picture. This means that children would have had time to subvocally name pictures, but that a significant naming time difference between short and long words of a few hundred milliseconds would be only detectable with more participants and/or more picture presentations per participant. Therefore, the lack of convergent results for children with a low memory span does not necessarily mean that their strategy self-reports were invalid.

For children with medium to high memory spans, naming was the most prevalent strategy for the majority of children and cumulative rehearsal was the dominant strategy for children with high spans. Furthermore, there was evidence that children were more likely to report and use cumulative rehearsal on the more difficult span+1 trials. It is important to note, again, that the self-report data as well as the indirect, behaviour-based indicators (recall performance and self-paced presentation times), support these conclusions. Even though we could not test whether each child's self-report correctly reflected their strategic behaviour on that trial, the convergent results of the multiple methods indicated that the strategy self-reports were valid at a group level for children with medium or high memory spans.

Only a minority of children maintained the same strategy throughout the experiment, the rest reporting two or more strategies. As discussed in 2.2.1 our efforts to facilitate self-reports in young children could alter task behaviour and encourage trying out several strategies. Because behavioural indicators with data aggregated across trials cannot really capture trial-to-trial variation in task behaviour or experimentation with strategies, we could not directly address this question. But as behavioural indicators and self-reports converge on the conclusion that naming and cumulative rehearsal were used, and as studies with think aloud procedures show that multiple forms of rehearsal are used not only across trials but also across the different serial positions within a trial (Lehmann & Hasselhorn, 2007; Tan & Ward, 2008), we see no strong reason to doubt that children would use various strategies even without being made aware of different rehearsal options by the self-report procedure. This mix of strategies and the link between list length or task difficulty and strategy choice (at least found in self-reports and in presentation times for high span children) suggests that the overlapping waves theory of cognitive development (Siegler, 1999, 2016) might be a better description of rehearsal development than a staircase view, i.e. that older but not younger children use rehearsal, or rehearsal being conceptualized as hard-wired maintenance mechanism. Siegler's theory posits that over prolonged periods, children use multiple ways of thinking and acting to solve similar tasks, so that previously dominant strategies are not immediately replaced, but new strategies are added. New effective strategies are usually preferred with less effective strategies gradually being phased out. Research on strategy use in free recall tasks (e.g. Lehmann & Hasselhorn, 2007), in word reading (e.g., Lindberg et al., 2014), or arithmetic strategies (e.g. Van der Ven, Boom, Kroesbergen, & Leseman, 2012) has supported the overlapping waves theory in primary school children, but this framework so far has not been used to conceptualise children's immediate serial recall.

Most of the intra-individual strategy variability from self-reports could not be explained by the different task conditions. However, strategy choice was related to the difficulty of the task, with somewhat less naming and somewhat more cumulative rehearsal on span+1 trials. Whether these influences on strategy preference would generalize to difficulty levels considerably above or below span, cannot be answered yet, as strategy data for immediate serial tasks are lacking.

Previous findings that most 10-year-olds used cumulative rehearsal while most 7-year-olds used naming (Henry et al., 2000) were supported by a clear age effect on cumulative rehearsal. But importantly, the current results qualify and expand the earlier findings. Firstly, analyses revealed that the likelihood of using cumulative rehearsal was linked to age as well as memory span level. This means that older children were more likely to rehearse cumulatively, even when span differences where controlled, and that young children with high spans were likely to report more cumulative rehearsal. Secondly, results showed that children cannot simply be classified as either using naming or cumulative rehearsal, although these two strategies were the dominant ones. Our finding of a gradual shift to more cumulative rehearsal, with age as well as higher memory span, is similar to the strategy use findings reported in free recall tasks (Lehmann & Hasselhorn, 2007). Similarities are not entirely surprising, considering that for adults, similar patterns of rehearsal according to rehearse-aloud protocols, as well as similar effects in experimental rehearsal indicators, can be found for free recall and immediate serial recall (Bhatarah, Ward, Smith, & Hayes, 2009). In serial recall in adults, the proportions of different forms of rehearsal change with presentation speed and with serial position during list presentation (Tan & Ward, 2008). Thus, it is unlikely that there is one mature adult default strategy like cumulative rehearsal. Again, different forms of rehearsal seem to coexist and might be applied depending on task demands as predicted by the overlapping waves theory (Siegler, 1999).

Our results showed that children's use of naming and cumulative rehearsal was linked to their baseline memory span (for a similar finding linking span level to 'no phonological recoding' vs. 'naming as a simple verbal strategy', see Poloczek et al., 2016). This relationship could be because cumulative rehearsal improves baseline recall (but see Lewandowsky & Oberauer, 2015) or because limits to memory capacity restrict the use of cumulative rehearsal (Cowan & Vergauwe, 2015; Jarrold & Hall, 2013). Baseline span was assessed with a medium presentation rate of one word every 2 seconds, this presentation rate could have allowed cumulative rehearsal and thereby affected baseline span. A further explanation could be, that the observed link between baseline memory span and cumulative rehearsal was caused by the design choice to titrate list length during the experiment to baseline span. Research varying the recall instruction in immediate recall tasks suggests that (adult) participants have privileged access to the first, the penultimate, and the final serial position (Tan, Ward, Paulauskaite, & Markou, 2016). All or most items for children with a span of 2 or 3 were in these three privileged positions. As cumulative rehearsal seems to be an optional maintenance strategy, it could be that children with low to medium spans chose not to use this strategy, because it was not necessary given the relatively short trials. Thus, future research could assess span with fast paced presentation to block rehearsal (Tan & Ward, 2008) and use multiple list lengths irrespective of memory span to shed light on whether there really is a link between span and cumulative rehearsal and how this link can be explained.

Another question is whether our findings transfer to immediate serial recall tasks with auditory presentation. For the current study, a visual presentation and non-verbal response format was crucial to avoid the possibility that word length effects were merely caused by verbal responses (Cowan et al., 1992). For future research with self-paced presentation time and self-reports as rehearsal indicators, this restriction would not apply, so rehearsal could be studied in the more common serial recall tasks with verbal presentation and recall.

5 Conclusions

The study demonstrated that combining behavioural indicators with self-reports provides novel insights into verbal rehearsal and naming in children between 6/7 and 10 years. The traditional view is that beyond the age of 7 years, adult-like cumulative rehearsal emerges and further developmental changes are caused by rehearsal processes becoming more effective due to increases in articulation rate (Gathercole, 1998). However, the methods that have led to this conclusion have been criticised (e.g. Jarrold and Hall, 2013). More importantly, the evidence provided in the current study suggests that it is more probable that children use a mix of rehearsal strategies during the primary school years, with naming and cumulative rehearsal as the dominant, but not exclusive strategies. Changes in the likelihood of using different forms of rehearsal were linked to task difficulty, age and memory span. Our research has provided a more nuanced picture of rehearsal use, especially in relation to inter-individual differences and intra-individual variability, and by a consideration of the possibility that the overlapping waves theory could be applied to immediate serial recall.

Disclosure of interest. The authors report no conflicts of interest.

References

Baddeley, A. D. (1986). *Working memory*. New York, NY: Oxford University Press.

Baddeley, A., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *36*, 233–252.

Belmont, J., & Butterfield, E. (1971). Learning strategies as determinants of memory deficiencies. *Cognitive Psychology*, *2*, 411–420.

Bhatarah, P., Ward, G., Smith, J., & Hayes, L. (2009). Examining the relationship between free recall and immediate serial recall: Similar patterns of rehearsal and similar effects of word length, presentation rate, and articulatory suppression. *Memory & Cognition*, *37*, 689–713.

Browne, W. J. (2017). *MCMC Estimation in MLwiN v3.00.* Centre for Multilevel Modelling, University of Bristol.

Camos, V. (2015). Storing verbal information in working memory. *Current Directions in Psychological Science*, *24*, 440–445.

Charlton, C., Rasbash, J., Browne, W. J., Healy, M., & Cameron, B. (2018). *MLwiN Version 3.02*. Centre for Multilevel Modelling, University of Bristol.

Cowan, N., Day, L., Saults, J. S., Keller, T. A., Johnson, T., & Flores, L. (1992). The role of verbal output time in the effects of word length on immediate memory. *Journal of Memory and Language*, *31*, 1–17.

Cowan, N., & Vergauwe, E. (2015). Applying how adults rehearse to understand how rehearsal may develop. *Frontiers in Psychology*, *5*, 1538.

Ferguson, A. N., Bowey, J. A., & Tilley, A. (2002). The association between auditory memory span and speech rate in children from kindergarten to sixth grade. *Journal of Experimental Child Psychology*, *81*, 141–56.

Flavell, J. H., Beach, D. R., & Chinsky, J. M. (1966). Spontaneous verbal rehearsal in a memory task as a function of age. *Child Development*, *37*, 283–299.

Flavell, J. H., Green, F. L., & Flavell, E. (2000). Development of children's awareness of their own thoughts. *Journal of Cognition and Development*, *1*, 97–112.

Gathercole, S. E. (1998). The development of memory. *Journal of Child Psychology and Psychiatry*, *39*, 3–27.

Henry, L. A. (1991). The effects of word length and phonemic similarity in young children's short-term memory. *The Quarterly Journal of Experimental Psychology Section A*, *43*, 35–52.

Henry, L. A. (2008). Short-term memory coding in children with intellectual disabilities. *American Journal on Mental Retardation, 113*, 187–200.

Henry, L. A., Messer, D., Luger-Klein, S., & Crane, L. (2012). Phonological, visual, and semantic coding strategies and children's short-term picture memory span. *Quarterly Journal of Experimental Psychology*, *65*, 2033–2053.

Henry, L. A., Turner, J. E., Smith, P. T., & Leather, C. (2000). Modality effects and the development of the word length effect in children. *Memory*, *8*, 1–17.

Hitch, G. J., Halliday, M. S., Dodd, A., & Littler, J. E. (1989). Development of rehearsal in short-term memory: Differences between pictorial and spoken stimuli. *British Journal of Developmental Psychology*, *7*, 347–362.

Hulme, C., Thomson, N., Muir, C., & Lawrence, A. (1984). Speech rate and the development of short-term memory span. *Journal of Experimental Child Psychology*, *38*, 241–253.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. Journal of Memory and Language, 59(4), 434–446.

Jarrold, C., & Hall, D. (2013). The development of rehearsal in verbal short-term memory. *Child Development Perspectives*, *7*, 182–186.

Jarrold, C., Danielsson, H., & Wang, X. (2015). Absolute and proportional measures of potential markers of rehearsal, and their implications for accounts of its development. *Frontiers in Psychology*, *6*, 299.

Lehmann, M., & Hasselhorn, M. (2007). Variable memory strategy use in children's adaptive intratask learning behavior: developmental changes and working memory influences in free recall. *Child Development*, *78*, 1068–1082.

Lewandowsky, S., & Oberauer, K. (2015). Rehearsal in serial recall: An unworkable solution to the nonexistent problem of decay. *Psychological Review*, *122*, 674–699.

Lindberg, S., Lonnemann, J., Linkersdörfer, J., Biermeyer, E., Mähler, C., Hasselhorn, M., & Lehmann, M. (2011). Early strategies of elementary school children's single word reading. *Journal of Neurolinguistics*, *24*, 556–570.

McGilly, K. & Siegler, R.S. (1989). How children choose among serial recall strategies. *Child Development, 60*, 172-182.

Morra, S. (2015). How do subvocal rehearsal and general attentional resources contribute to verbal short-term memory span? *Frontiers in Psychology*, *6*, 145.

Poloczek, S., Henry, L. A., Danielson, H., Büttner, G., Mähler, C., Messer, D. J., … Van der Molen, M. J. (2016). Strategic verbal rehearsal in adolescents with mild intellectual disabilities: A multi-centre European study. *Research in Developmental Disabilities*, *58*, 83–93.

Raven, J. C., & Horn, R. (2009). *Ravens Progressive Matrices and Vocabulary Scales. Deutsche Bearbeitung und Normierung*. [German adaptation and standardization]. Frankfurt am Main: Pearson.

Siegler, R. S. (1999). Strategic development. *Trends in Cognitive Sciences*, *3*, 430–435.

Siegler, R. S. (2016). Continuity and change in the field of cognitive development and in the perspectives of one cognitive developmentalist. *Child Development Perspectives*, *10*, 128–133.

Tan, L., & Ward, G. (2008). Rehearsal in immediate serial recall. *Psychonomic Bulletin & Review*, *15*, 535–542.

Tan, L., Ward, G., Paulauskaite, L., & Markou, M. (2016). Beginning at the beginning: Recall order and the number of words to be recalled. *Journal of Experimental Psychology: Learning Memory and Cognition*, *42*(8), 1282–1292.

Van der Ven, S. H. G., Boom, J., Kroesbergen, E. H., & Leseman, P. P. M. (2012). Microgenetic patterns of children's multiplication learning: confirming the overlapping waves model by latent growth modeling. *Journal of Experimental Child Psychology*, *113*, 1–19.

Zhang, Z., Parker, R. M. A., Charlton, C. M. J., Leckie, G., & Browne, W. J. (2016). R2MLwiN: A Package to Run MLwiN from within R. *Journal of Statistical Software, 72(10)*, 1-43.

Table 1. Number of children classified by age (rows) and levels of picture span at baseline testing (columns).

Age	Span					*Total*
	2	3	4	5	6	
6	0	7	2	0	0	*9*
7	7	23	6	1	0	*37*
8	4	16	9	3	1	*33*
9	2	7	7	3	0	*19*
10	0	2	7	0	0	*9*
11	0	0	1	0	0	*1*
Total	*13*	*55*	*32*	*7*	*1*	*108*

Table 2. Proportion of pictures recalled correctly in their position for all experimental conditions of the session without self-reports for children grouped according to their baseline memory span

Children with a	List Length	Pacing	Short		Long	
			M	*(SD)*	*M*	*(SD)*
span of 2	span	fixed	.94	(.11)	.95	(.11)
(*n*=13)		self	.93	(.13)	.91	(.11)
	span+1	fixed	.85	(.17)	.80	(.22)
		self	.78	(.21)	.69	(.30)
span of 3	span	fixed	.91	(.17)	.86	(.18)
(*n*=55)		self	.89	(.18)	.82	(.21)
	span+1	fixed	.68	(.24)	.54	(.25)
		self	.67	(.30)	.56	(.27)
span of 4	span	fixed	.87	(.17)	.72	(.21)
or higher		self	.88	(.15)	.71	(.22)
(*n*=40)	span+1	fixed	.62	(.22)	.43	(.22)
		self	.67	(.26)	.46	(.22)

Note. SD is the between-participant standard deviation after within-participant aggregation of trials belonging to the same condition.

Table 3. Logistic GLMM to test differences in likelihood of recalling picture in correct position in the session without self-reports

Fixed Part	**β**	**95% CI**	***p***
Intercept	1.46	[1.27, 1.66]	<.001
Word length (short = -1, long = +1)	**-0.31**	**[-0.38, -0.24]**	**<.001**
List length (span = -1, span+1 = +1)	**-0.81**	**[-0.88, -0.74]**	**<.001**
Pacing (fixed = -1, self-paced = +1)	**-0.07**	**[-0.15, 0.00]**	**.05**
Memory Span (span of 2 = -1, span of 3 = 0, span >=4 = +1)	**-0.60**	**[-0.89, -0.32]**	**<.001**
Word Length x List Length	-0.01	[-0.08, 0.06]	.69
Word Length x Pacing	-0.02	[-0.10, 0.05]	.52
Word Length x Memory span	**-0.23**	**[-0.33, -0.13]**	**<.001**
List Length x Pacing	0.04	[-0.03, 0.11]	.22
List Length x Memory span	**0.10**	**[0.00, 0.19]**	**.05**
Pacing x Memory span	**0.13**	**[0.03, 0.22]**	**.008**
Word Length x List Length x Pacing	0.03	[-0.04, 0.11]	.37
Word Length x List Length x Memory span	**0.09**	**[0.00, 0.19]**	**.05**
List Length x Pacing x Memory span	0.00	[-0.10, 0.09]	.94
Word Length x Pacing x Memory span	-0.01	[-0.10, 0.08]	.84
Word Length x List Length x Pacing x Memory span	-0.01	[-0.11, 0.08]	.81
Random Part	***u***	**95% CI**	**Δ *DIC***
Participant Intercept	0.82	[0.60, 1.10]	1082.0

Note. CI = credible interval, the interval containing the middle 95% of estimates from the MCMC estimation; Δ DIC = Improvement in Deviance Information Criterion, when random effect added to model

Table 4. Self-paced presentation times (in ms) at each serial position in all experimental conditions of the session without self-reports for children grouped according to their baseline memory span.

Children with a	Difficulty	Word length	Position 1		Position 2		Position 3		Position 4		Position 5	
			M	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*
span of 2	span	short	2674	(1262)	1748	(620)						
(*n*=13)		long	2652	(1157)	1787	(641)						
	span+1	short	2208	(840)	1559	(497)	1711	(681)				
		long	2225	(635)	1728	(645)	1379	(546)				
span of 3	span	short	2328	(872)	1756	(643)	1562	(691)				
(*n*=55)		long	2624	(1108)	1873	(766)	1516	(665)				
	span+1	short	2332	(787)	1817	(845)	1757	(811)	1923	(1235)		
		long	2569	(1156)	1816	(704)	1946	(1094)	1989	(1204)		
span of 4	span	short	2799	(1310)	2375	(1632)	2471	(1481)	2281	(1415)		
or higher		long	2671	(1277)	2327	(950)	2531	(1430)	2702	(1665)		
(*n*=40)	span+1	short	2659	(1089)	2126	(1043)	2348	(1249)	2453	(1747)	2719	(1763)
		long	2867	(1426)	2458	(1226)	2554	(1316)	2840	(1937)	3177	(2193)

Note. SD is the between-participant standard deviation

Table 5: LMM for self-paced presentation times of the session without self-reports, from position 2 onwards due to delayed response for picture 1

Fixed Part	**β**	**95% CI**	***p***
Intercept	1.89	[1.71, 2.06]	<.001
Word length (short = -1, long = +1)	0.04	[-0.02, 0.11]	0.19
List length (span = -1, span+1 = +1)	-0.03	[-0.09, 0.03]	0.34
Position (Pos2 = 0, Pos3 = 1, Pos4 = 2 …)	-0.11	[-0.23, 0.00]	.06
Memory Span (span of 2 = -1, span of 3 = 0, span >=4 = +1)	**0.37**	**[0.12, 0.62]**	**.004**
Word Length x List Length	0.01	[-0.05, 0.07]	.77
Word Length x Position	-0.04	[-0.12, 0.05]	.40
Word Length x Memory span	0.03	[-0.05, 0.11]	.50
List Length x Position	**0.18**	**[0.09, 0.27]**	**<.001**
List Length x Memory span	-0.01	[-0.1, 0.07]	.77
Position x Memory span	**0.25**	**[0.11, 0.40]**	**<.001**
Word Length x List Length x Position	0.02	[-0.06, 0.11]	.61
Word Length x List Length x Memory span	0.06	[-0.02, 0.15]	.13
List Length x Position x Memory span	**-0.12**	**[-0.21, -0.02]**	**.02**
Word Length x Position x Memory span	0.07	[-0.02, 0.17]	.11
Word Length x List Length x Position x Memory span	-0.06	[-0.15, 0.03]	.19
Random Part	***u***	**95% CI**	**Δ *DIC***
Participant Intercept	0.68	[0.50, 0.91]	1753.0
Covariance Intercept - Slope	0.01	[-0.05, 0.08]	212.6
Position Slope	0.11	[0.07, 0.16]	
Residual Variance	1.23	[1.17, 1.29]	

Note. CI = credible interval; Δ DIC = Improvement in Deviance Information Criterion

Table 6. Percentage of children who reported using the different strategies.

self-reported strategy use	never (0%)	rarely (4 - 25%)	sometimes (26 - 50%)	often (51 - 75%)	mostly (76 - 96%)	always (100%)
naming	12.0%	35.2%	27.8%	7.4%	7.4%	10.2%
complete rehearsal	46.3%	49.1%	4.6%	0.0%	0.0%	0.0%
cumulative rehearsal	13.0%	38.9%	21.3%	12.0%	7.4%	7.4%
visual strategy	42.6%	50.0%	5.6%	0.9%	0.9%	0.0%
no strategy	47.2%	41.7%	6.5%	1.9%	2.8%	0.0%

Note. Self-reports were assessed on a trial-by-trial basis. Children were assigned to the frequency categories based on their 24 answers.

Table 7. Percent use of strategies for children grouped by their baseline picture span.

Group	Naming	Complete Rehearsal	Cumulative Rehearsal	Visual Strategy	No Strategy
Span of 2 (n_{ch}=13, n_{tr}=310)	42.3%	8.4%	18.1%	14.5%	16.8%
Span of 3 (n_{ch}=55, n_{tr}=1314)	46.7%	8.8%	21.2%	10.9%	12.6%
Span of >=4 (n_{ch}=40, n_{tr}=958)	23.2%	5.3%	59.0%	6.5%	6.1%
All children (N_{ch}=108, N_{tr}=2582)	37.4%	7.4%	34.8%	9.7%	10.7%

Note. N_{ch} = number of children; N_{tr} = number of trials, in 10 trials (0.4% of all trials) self-reports were missing, these trials were excluded from the calculations for the table and from the GLMM analyses.

Table 8. Results of the five final logistic GLMMs predicting whether a certain strategy option was reported or not.

	Naming			Complete Rehearsal			Cumulative Rehearsal			Visual Strategy			No Strategy		
Fixed Part	**β**	**95% CI**	***p***	**β**	**95% CI**	***p***	**β**	**95% CI**	***p***	**β**	**95% CI**	***p***	**β**	**95% CI**	***p***
Intercept	-0.45	[-1.13, 0.24]	.19	-2.78	[-3.33, -2.30]	<.001	-1.86	[-2.46, -1.28]	<.001	-2.66	[-3.24, -2.13]	<.001	-2.96	[-3.77, -2.25]	<.001
Level 1 *(trials)*															
Word length (short = -1, long = +1)	-0.04	[-0.15, 0.07]	.45	0.10	[-0.06, 0.25]	.21	-0.05	[-0.16, 0.05]	.33	0.10	[-0.04, 0.24]	.18	0.11	[-0.04, 0.26]	.16
List length (span = -1, span+1 = +1)	**-0.21**	**[-0.32, -0.10]**	**<.001**	-0.03	[-0.19, 0.13]	.74	**0.25**	**[0.14, 0.36]**	**<.001**	-0.02	[-0.16, 0.13]	.83	0.03	[-0.12, 0.18]	.67
Pacing (fixed = -1, self-paced = +1)	-0.05	[-0.16, 0.05]	.32	-0.09	[-0.25, 0.07]	.27	0.10	[-0.01, 0.21]	.07	-0.08	[-0.22, 0.06]	.28	-0.09	[-0.24, 0.06]	.22
Level 2 *(children)*															
Span = 2	0.07	[-1.44, 1.57]	.93	0.13	[-0.93, 1.18]	.81	-0.27	[-1.49, 0.97]	.67	0.47	[-0.65, 1.59]	.42	0.71	[-0.78, 2.22]	.35
Span >= 4	**-1.94**	**[-3.03, -0.88]**	**<.001**	-0.18	[-1.03, 0.64]	.67	**2.16**	**[1.32, 3.10]**	**<.001**	-0.44	[-1.37, 0.44]	.33	-0.61	[-1.86, 0.61]	.33
Age (centered at 8.0 years)	-0.34	[-0.81, 0.11]	.14	**-0.55**	**[-0.92, -0.21]**	**.002**	**0.72**	**[0.33, 1.11]**	**<.001**	-0.23	[-0.63, 0.15]	.25	-0.47	[-1.00, 0.02]	.07
Nonverbal IQ (z-standardized)	**0.71**	**[0.31, 1.13]**	**<.001**	**-0.57**	**[-0.92, -0.25]**	**<.001**	-0.16	[-0.52, 0.18]	.36	**-0.41**	**[-0.76, -0.06]**	**.02**	-0.42	[-0.92, 0.05]	.09
Gender (boys = -1, girls = +1)	-0.05	[-0.51, 0.42]	.83	0.16	[-0.20, 0.52]	.37	0.01	[-0.38, 0.39]	.96	-0.02	[-0.40, 0.37]	.92	-0.07	[-0.58, 0.45]	.80
Random Part	***u***	**95% CI**	**Δ *DIC***	***u***	**95% CI**	**Δ *DIC***	***u***	**95% CI**	**Δ *DIC***	***u***	**95% CI**	**Δ *DIC***	***u***	**95% CI**	**Δ *DIC***
Participant Intercept	4.97	[3.32, 7.17]	933.9	1.87	[0.99, 3.23]	134.9	3.37	[2.23, 4.97]	640.5	2.60	[1.52, 4.18]	440.8	4.73	[2.80, 7.53]	255.5

Note. CI = Credible Interval; Δ DIC = Improvement in Deviance Information Criterion

Figure 1. Picture sequences to explain a) naming, b) complete rehearsal, c) cumulative rehearsal, d) visual strategy and e) no strategy